

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Исследование операций и принятие решений в задачах оптимизации,
управления и экономики

Плоткин Артем Владимирович

ВАРИАНТЫ МЕТОДА СОПРЯЖЕННЫХ ГРАДИЕНТОВ БЕЗ ТОЧНОГО
ЛИНЕЙНОГО ПОИСКА

Бакалаврская работа

Научный руководитель:

д. ф.-м. н., профессор В. Н. Малоземов

Рецензент:

к. ф.-м. н., научный сотрудник

М. В. Долгополик

Санкт-Петербург

2016

Saint Petersburg State University
Applied Mathematics and Computer Science
Operation Research and Decision Making in Optimisation,
Control and Economics Problems

Plotkin Artem Vladimirovich

VARIANTS OF CONJUGATE GRADIENT METHOD WITHOUT EXACT LINEAR
SEARCH

Bachelor's Thesis

Scientific Supervisor:

Professor V. N. Malozemov

Reviewer:

Research fellow M. V. Dolgopolik

Saint Petersburg

2016

Оглавление

Введение	4
Глава 1. О методе сопряженных градиентов	5
1.1. Минимизация квадратичной функции	5
1.2. Общая задача безусловной минимизации	6
1.3. Обновление	6
1.4. Линейный поиск	7
1.5. Параметр сопряженности	10
Глава 2. Сходимость	12
2.1. Условия	12
2.2. Теорема Зойтендейка	12
Глава 3. Численные эксперименты и их анализ	14
3.1. Целевая функция	14
3.2. Численные эксперименты	15
3.3. Анализ результатов	18
Глава 4. Минимизация циклической функции	20
4.1. Неравенство Шапиро	20
4.2. Постановка задачи и метод решения	20
4.3. Начальное приближение	21
4.4. Результаты вычислений	22
Глава 5. Геометрический вариант метода сопряженных градиентов . .	25
5.1. Особенности геометрического варианта	25
5.2. Сходимость	26
Заключение	29
Список литературы	30

Введение

Метод сопряженных градиентов применяется для решения задач безусловной минимизации. Метод имеет длительную историю развития, начальный этап которого был подробно описан Голубом и О'Лири в обзорной статье [10].

Первая работа в данной области принадлежит Хестенсу и Штифелю и датируется 1952 годом [1]. В этой работе метод сопряженных градиентов применялся для решение систем линейных уравнений. В 1964-м году Флетчер и Ривз [9] впервые использовали метод сопряженных градиентов для минимизации гладких функций.

Развитие метода продолжается и сейчас: разрабатываются новые варианты метода, исследуются вопросы сходимости.

Достоинствами метода сопряженных градиентов являются его простота и низкие затраты памяти, что делает его особенно эффективным при решении задач большой размерности.

В данной работе представлен обзор и сравнение различных вариантов метода сопряженных градиентов. Большое внимание уделяется наиболее эффективным способам осуществления линейного поиска. Исследуются вопросы сходимости. На примере задачи о минимизации циклической функции проверяется эффективность работы метода.

Глава 1

О методе сопряженных градиентов

1.1. Минимизация квадратичной функции

Рассмотрим задачу минимизации квадратичной функции

$$Q(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle \rightarrow \min_{x \in \mathbb{R}^n}, \quad (1.1)$$

где матрица A симметрична и положительно определена. Функция Q строго выпукла на \mathbb{R}^n , и условие $Q'(x) = 0$ служит критерием оптимальности для поставленной задачи.

В 1952-м году Хестенс и Штифель предложили эффективный метод решения задачи (1.1), который получил название *метод сопряженных градиентов* [1]. Опишем вычислительную схему данного метода.

Нулевой шаг. Возьмем произвольное начальное приближение $x_1 \in \mathbb{R}^n$ и вычислим значение градиента $g_1 = Q'(x_1) = Ax_1 - b$. Если $g_1 = \mathbb{O}$, то x_1 — решение задачи (1.1).

В противном случае задаем направление $d_1 = -g_1$.

k -й шаг. Пусть имеются x_k , d_k и $g_k \neq \mathbb{O}$. Вычислим следующее приближение x_{k+1} по формуле

$$\alpha_k = \frac{\langle g_k, g_k \rangle}{\langle Ad_k, d_k \rangle}, \quad (1.2)$$

$$x_{k+1} = x_k + \alpha_k d_k. \quad (1.3)$$

Далее найдем значение градиента $g_{k+1} = Q'(x_{k+1}) = g_k + \alpha_k Ad_k$. Если $g_{k+1} = \mathbb{O}$, то x_{k+1} — решение задачи (1.1). В противном случае вычисляем следующее направление d_{k+1} по формуле

$$\beta_k = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \quad (1.4)$$

$$d_{k+1} = -g_{k+1} + \beta_k d_k. \quad (1.5)$$

Не позднее $k = n$ метод найдет решение поставленной задачи. Коэффициент α_k обладает экстремальным свойством: минимум функции $Q(x_k + \alpha d_k)$ достигается при $\alpha = \alpha_k$. Данный факт позволяет сформулировать вычислительную схему метода без использования матрицы A .

Более подробный обзор метода сопряженных градиентов для решения задачи (1.1) можно найти в [2].

1.2. Общая задача безусловной минимизации

Предположение о том, что любая гладкая функция f в окрестности локального минимума хорошо аппроксимируется выпуклой квадратичной, привело к идеи использования метода сопряженных градиентов для решения общей задачи безусловной минимизации

$$f(x) \rightarrow \min_{x \in \mathbb{R}^n}, \quad (1.6)$$

где функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ непрерывно дифференцируема и ограничена снизу. Градиент целевой функции будем обозначать через $g(x)$.

Опишем вычислительную схему метода сопряженных градиентов для решения задачи (1.6).

Нулевой шаг. Возьмем произвольное начальное приближение $x_1 \in \mathbb{R}^n$ и вычислим значение градиента $g_1 = g(x_1)$. Если $g_1 = \mathbb{O}$, то x_1 — решение задачи (1.6). В противном случае задаем направление $d_1 = -g_1$.

k -й шаг. Пусть имеются x_k, d_k . Вычислим следующее приближение x_{k+1} по формуле

$$x_{k+1} = x_k + \alpha_k d_k, \quad (1.7)$$

где α_k — приближенное решение задачи одномерной минимизации

$$f(x_k + \alpha d_k) \rightarrow \min_{\alpha > 0}. \quad (1.8)$$

Далее найдем значение градиента $g_{k+1} = g(x_{k+1})$. Если $g_{k+1} = \mathbb{O}$, то x_{k+1} — решение задачи (1.1). В противном случае вычисляем следующее направление d_{k+1} по формуле

$$d_{k+1} = -g_{k+1} + \beta_k d_k. \quad (1.9)$$

Коэффициент β_k называется *параметром сопряженности*. Каждый вариант метода сопряженных градиентов имеет свое определение этого параметра. Однако, в случае, когда f — строго выпуклая квадратичная функция, а линейный поиск, решающий задачу (1.8), находит точный минимум, все варианты параметра β_k эквивалентны.

1.3. Обновление

Будем говорить, что выполняется *свойство убывания*, если

$$\langle g_k, d_k \rangle < 0. \quad (1.10)$$

Если данное свойство нарушается, то может и не существовать $\alpha > 0$ такого, что $f(x_k + \alpha d_k) < f(x_k)$. Чтобы бороться с этим явлением, метод сопряженных градиентов время от времени обновляют: берут текущее приближение x_k в качестве начального. На практике обновление делают каждые n итераций или при нарушении свойства убывания.

1.4. Линейный поиск

Рассмотрим подробнее задачу одномерной минимизации (1.8) и различные варианты линейного поиска для ее решения. Будем предполагать, что d_k является направлением убывания. Введем функцию $\varphi(\alpha) = f(x_k + \alpha d_k)$ и перепишем задачу (1.8) в упрощенном виде:

$$\varphi(\alpha) \rightarrow \min_{\alpha > 0}. \quad (1.11)$$

Отметим, что $\varphi'(\alpha) = \langle g(x_k + \alpha d_k), d_k \rangle$. Перейдем к описанию различных вариантов решения этой задачи.

Точный поиск. Наилучшим выбором для значения α_k является первая стационарная точка функции $\varphi(\alpha)$:

$$\alpha_k = \min\{\alpha > 0 \mid \varphi'(\alpha) = 0\}. \quad (1.12)$$

Тогда выполняется условие

$$\langle g_{k+1}, d_k \rangle = 0. \quad (1.13)$$

Из формулы (1.9) и условия (1.13) следует, что при таком выборе α_k условие убывания выполняется для всех $k \geq 1$. Однако использование такого поиска очень затратно и редко применяется на практике. Кроме того, условий гладкости и ограниченности целевой функции снизу недостаточно для существования стационарной точки функции $\varphi(\alpha)$. На практике применяется линейный поиск, основывающийся на выборе α_k , которое удовлетворяет некоторым условиям. Перечислим наиболее часто используемые из них.

Условие Армихо.

$$\varphi(\alpha) \leq \varphi(0) + \mu \alpha \langle g_k, d_k \rangle, \quad (1.14)$$

где $0 < \mu < 1$ — фиксированный параметр.

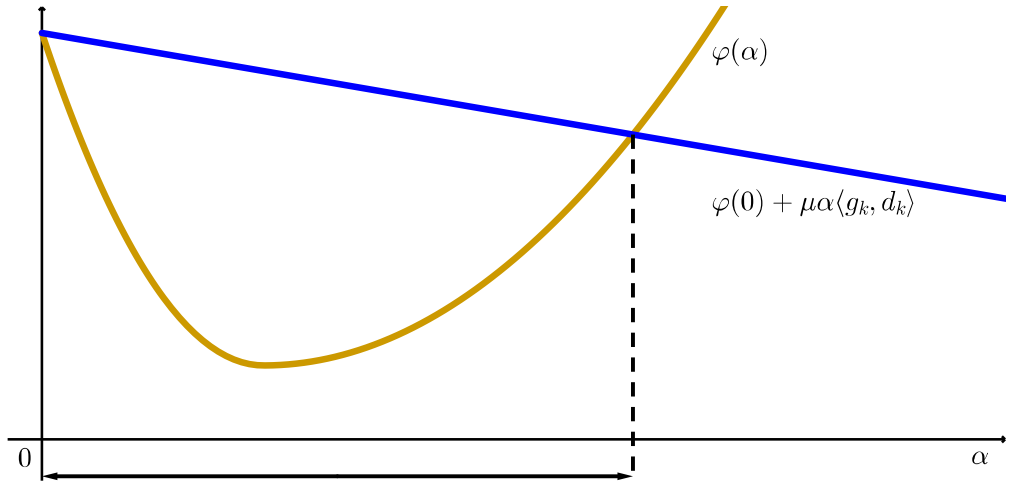


Рис. 1.1. Условие Армихо.

Условие Гольдштейна.

$$\varphi(0) + (1 - \mu)\alpha\langle g_k, d_k \rangle \leq \varphi(\alpha) \leq \varphi(0) + \mu\alpha\langle g_k, d_k \rangle, \quad (1.15)$$

где $0 < \mu < \frac{1}{2}$ — фиксированный параметр.

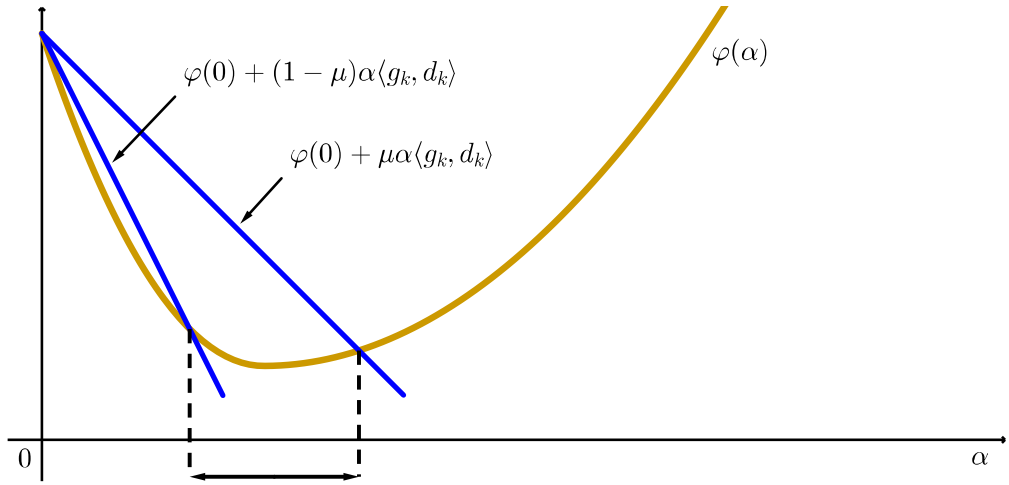


Рис. 1.2. Условие Гольдштейна.

Стандартные условия Вулфа.

$$\varphi(\alpha) \leq \varphi(0) + \mu\alpha\langle g_k, d_k \rangle, \quad (1.16)$$

$$\varphi'(\alpha) \geq \nu\langle g_k, d_k \rangle, \quad (1.17)$$

где $0 < \mu < \nu < 1$ — фиксированные параметры.

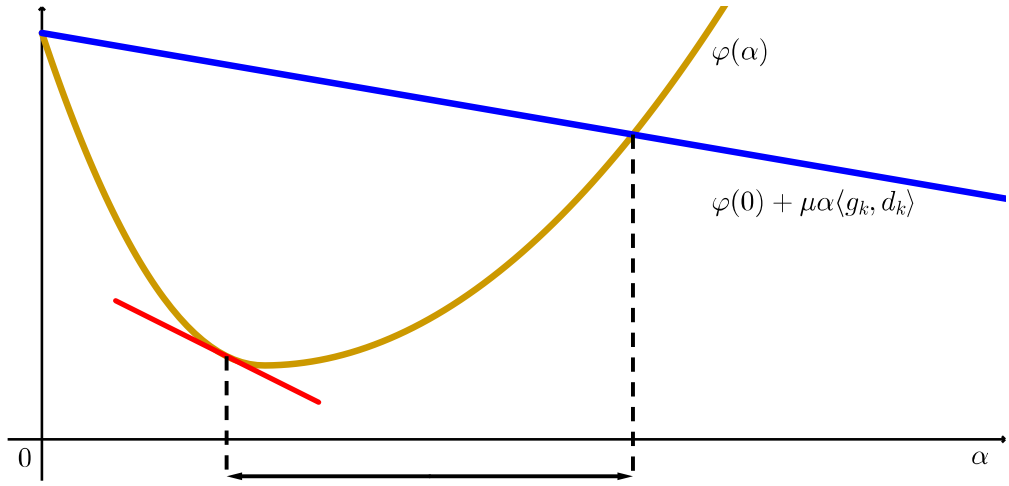


Рис. 1.3. Стандартные условия Вулфа.

Усиленные условия Вулфа.

$$\varphi(\alpha) \leq \varphi(0) + \mu\alpha\langle g_k, d_k \rangle, \quad (1.18)$$

$$|\varphi'(\alpha)| \leq -\nu\langle g_k, d_k \rangle, \quad (1.19)$$

где $0 < \mu < \nu < 1$ — фиксированные параметры.

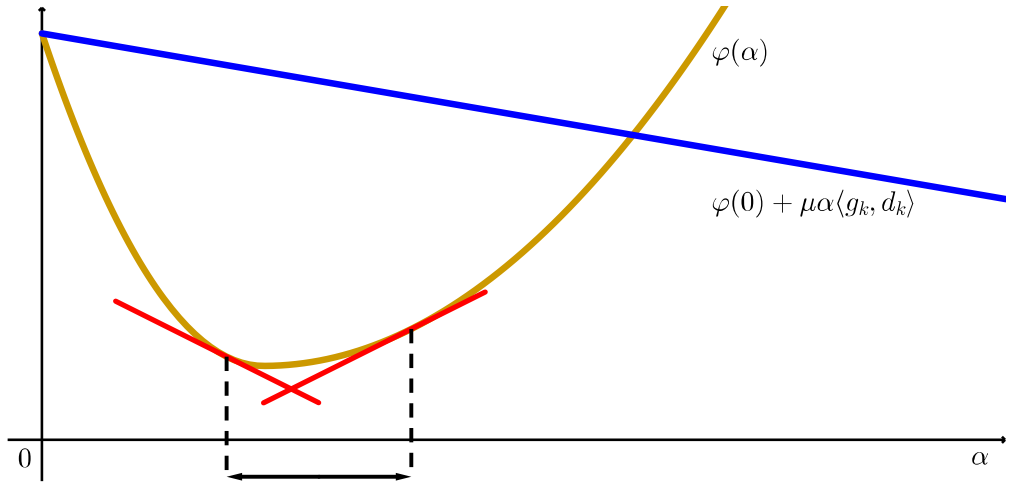


Рис. 1.4. Усиленные условия Вулфа.

Теорема 1. Пусть функция f непрерывно дифференцируема и ограничена снизу на \mathbb{R}^n , и выполняется условие убывания. Тогда существует непустой интервал α , удовлетворяющий стандартным и усиленным условиям Вулфа.

Доказательство. Рассмотрим луч $l(\alpha) = \varphi(0) + \mu\alpha\langle g_k, d_k \rangle, \alpha > 0$. Такой луч пересекает график функции $\varphi(\alpha)$. Пусть α' является абсциссой первой точки пересечения. Тогда

выполняется

$$\varphi(\alpha') = \varphi(0) + \mu\alpha\langle g_k, d_k \rangle. \quad (1.20)$$

По теореме Лагранжа о среднем значении существует $\alpha'' \in (0, \alpha')$, при котором выполняется

$$\varphi(\alpha') = \varphi(0) + \alpha'\varphi'(\alpha''). \quad (1.21)$$

Тогда

$$\varphi'(\alpha'') = \mu\langle g_k, d_k \rangle > \nu\langle g_k, d_k \rangle. \quad (1.22)$$

Получили, что для α'' верно

$$\varphi(\alpha'') < \varphi(0) + \mu\alpha\langle g_d, d_k \rangle, \quad (1.23)$$

$$\nu\langle g_k, d_k \rangle < \varphi(\alpha'') < 0. \quad (1.24)$$

Из полученных неравенств и гладкости функции f следует выполнение стандартных и усиленных условий Вулфа в некоторой окрестности α'' . \square

Для условий Армихо и Гольдштейна такой результат очевиден. Подробности реализации описанных выше вариантов линейного поиска можно найти в [3].

1.5. Параметр сопряженности

Введем обозначение $y_k = g_{k+1} - g_k$ и перечислим различные варианты выбора параметра сопряженности β_k в порядке их появления:

$\beta_k^{HS} = \frac{\langle g_{k+1}, y_k \rangle}{\langle d_k, y_k \rangle}$	Хестенс, Штифель	(1952 г.)
$\beta_k^{FR} = \frac{ g_{k+1} ^2}{ g_k ^2}$	Флетчер, Ривз	(1964 г.)
$\beta_k^{PR} = \frac{\langle g_{k+1}, y_k \rangle}{ g_k ^2}$	Полак, Рибьер	(1969 г.)
$\beta_k^{CD} = \frac{ g_{k+1} ^2}{\langle -d_k, g_k \rangle}$	Флетчер	(1987 г.)
$\beta_k^{LS} = \frac{\langle g_{k+1}, y_k \rangle}{\langle -d_k, g_k \rangle}$	Лиу, Стори	(1991 г.)
$\beta_k^{DY} = \frac{ g_{k+1} ^2}{\langle d_k, y_k \rangle}$	Дай, Юань	(1999 г.)
$\beta_k^{HZ} = \left\langle y_k - 2d_k \frac{ y_k ^2}{\langle d_k, y_k \rangle}, \frac{g_{k+1}}{\langle d_k, y_k \rangle} \right\rangle$	Хагер, Жанг	(2005 г.)

Таблица 1.1. Варианты параметра β_k .

Глава 2

Сходимость

2.1. Условия

Анализ сходимости метода сопряженных градиентов осуществляется при более строгих условиях на целевую функцию:

Условие 1. Множество уровня $\mathcal{L} = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_1)\}$, где x_1 — начальное приближение, ограничено.

Условие 2. В некоторой окрестности \mathcal{N} множества \mathcal{L} функция f непрерывно дифференцируема, и ее градиент $g(x)$ удовлетворяет условию Липшица: существует константа $L > 0$ такая, что

$$\|g(x) - g(y)\| \leq L\|x - y\| \quad \forall x, y \in \mathcal{N}.$$

Данных условий достаточно, чтобы гарантировать существования стационарной точки функции $\varphi(\alpha)$, что позволяет использовать точный линейный поиск (1.12).

Будем говорить, что метод сопряженных градиентов сходится, если выполняется условие

$$\lim_{k \rightarrow \infty} \|g_k\| = 0. \quad (2.1)$$

2.2. Теорема Зойтендейка

Основным инструментом в доказательстве сходимости того или иного варианта метода сопряженных градиентов является теорема Зойтендейка. Для ее формулировки нам потребуется ввести угол Θ_k между $-g_k$ и d_k :

$$\cos \Theta_k = \frac{\langle -g_k, d_k \rangle}{\|g_k\| \|d_k\|}. \quad (2.2)$$

Теорема 2 (Зойтендейк). Пусть функция f удовлетворяет Условиям 1, 2. Рассмотрим итеративный метод вида (1.7). Пусть условие убывания выполняется для всех $k \geq 1$, а значение α_k удовлетворяет стандартным условиям Вулфа (1.16), (1.17). Тогда существует константа $c > 0$ такая, что

$$f(x_k) - f(x_{k+1}) \geq c \cos^2 \Theta_k \|g_k\|^2 \quad \forall k \geq 1. \quad (2.3)$$

Доказательство. Из условия (1.17) получаем

$$\langle g_{k+1} - g_k, d_k \rangle \geq (\nu - 1) \langle g_k, d_k \rangle. \quad (2.4)$$

Из Условия 2 также имеем

$$\langle g_{k+1} - g_k, d_k \rangle \leq \alpha_k L \|d_k\|^2. \quad (2.5)$$

Комбинируя эти два неравенства, получим

$$\alpha_k \geq \frac{\nu - 1}{L} \langle g_k, d_k \rangle, \quad (2.6)$$

что совместно с условием (1.16) приводит к

$$f(x_k) - f(x_{k+1}) \geq \frac{\mu(1 - \nu)}{L} \frac{\langle g_k, d_k \rangle^2}{\|d_k\|^2}. \quad (2.7)$$

Используя выражение (2.2), перепишем этот результат в виде

$$f(x_k) - f(x_{k+1}) \geq c \cos^2 \Theta_k \|g_k\|^2, \quad (2.8)$$

где $c = \mu(1 - \nu)/L > 0$. □

Следствие 1. Пусть выполняются предположения Теоремы 2. Тогда

$$\sum_{i=1}^{\infty} \cos^2 \Theta_k \|g_k\|^2 < \infty. \quad (2.9)$$

Доказательство. Верность (2.9) следует из неравенства (2.3) и ограниченности функции f снизу. □

Замечание 1. Вывод теоремы Зойтендейка остается справедливым, если α_k вычисляется точным поиском (1.12).

Доказательство. Действительно, пусть $x_{k+1} = x_k + \alpha_k d_k$, где α_k находится точным поиском (1.12). Если α_k удовлетворяет стандартным условиям Вулфа, то неравенство (2.8) выполняется. Если же неравенство (1.16) нарушается, то найдется $\alpha' < \alpha_k$, удовлетворяющее стандартным условиям Вулфа, для которого верно неравенство (2.8). Но $f(x_{k+1}) < f(x_k + \alpha' d_k)$, а значит

$$f(x_k) - f(x_{k+1}) \geq f(x_k) - f(x_k + \alpha' d_k) \geq c \cos^2 \Theta_k \|g_k\|^2. \quad (2.10)$$

□

Доказательство сходимости различных вариантов метода сопряженных градиентов можно найти в [4], [11], [12].

Глава 3

Численные эксперименты и их анализ

3.1. Целевая функция

Рассмотрим на примере функции Стайблински-Танга поведение различных вариантов метода сопряженных градиентов. Функция задается выражением

$$f(x) = \frac{1}{2} \sum_{i=1}^n x_i^4 - 16x_i^2 + 5x_i. \quad (3.1)$$

Глобальным минимум достигается в точке (η, \dots, η) , где $\eta \approx -2.904$. Использовать данную функцию будем при $n = 2$, чтобы наглядно проиллюстрировать результаты.

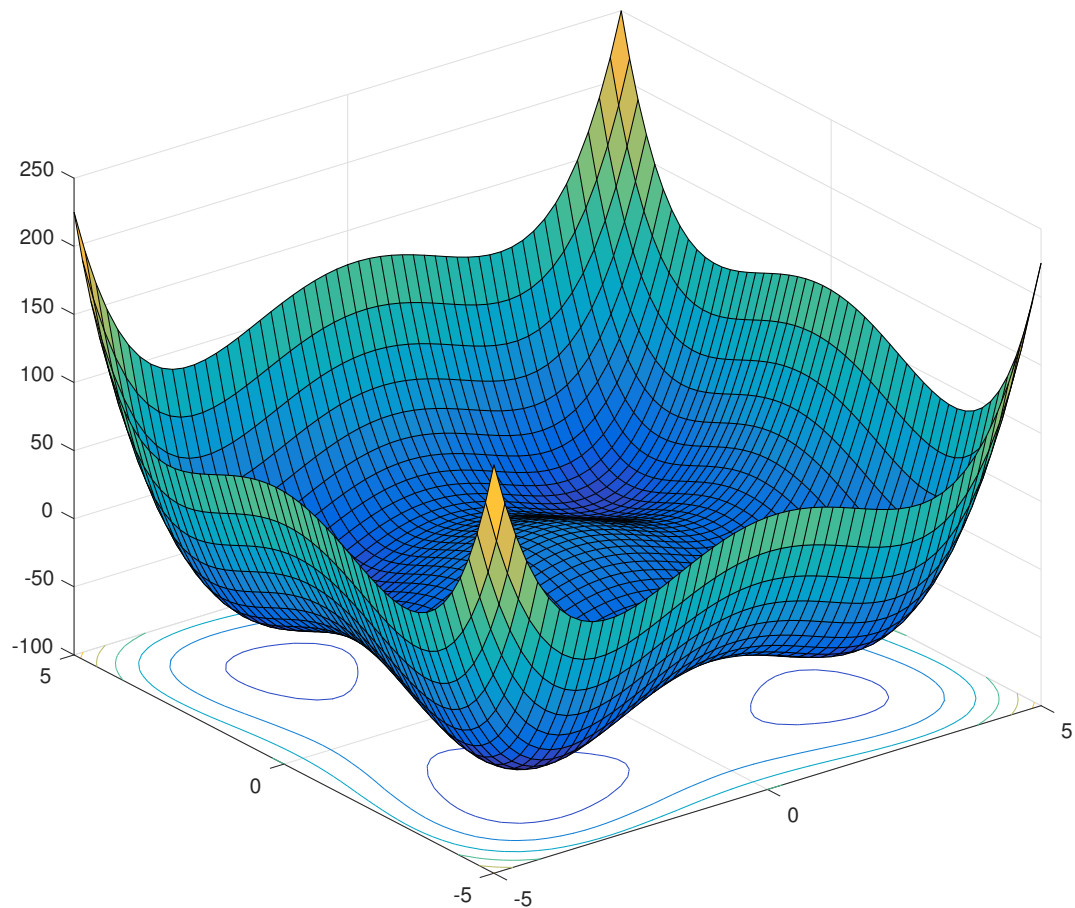


Рис. 3.1. Функция Стайблински-Танга при $n = 2$.

3.2. Численные эксперименты

Сравнивались методы Флетчера-Ривза, Хестенса-Штифеля, Полака-Рибьера и Хайера-Жанга. Линейный поиск удовлетворял усиленным условиям Вульфа с параметрами $\mu = 0.4$, $\nu = 0.6$. Обновление методов происходило только при нарушении свойства убывания. Метод останавливался на итерации k , если выполнялось условие $\|g_k\| < 10^{-6}$.

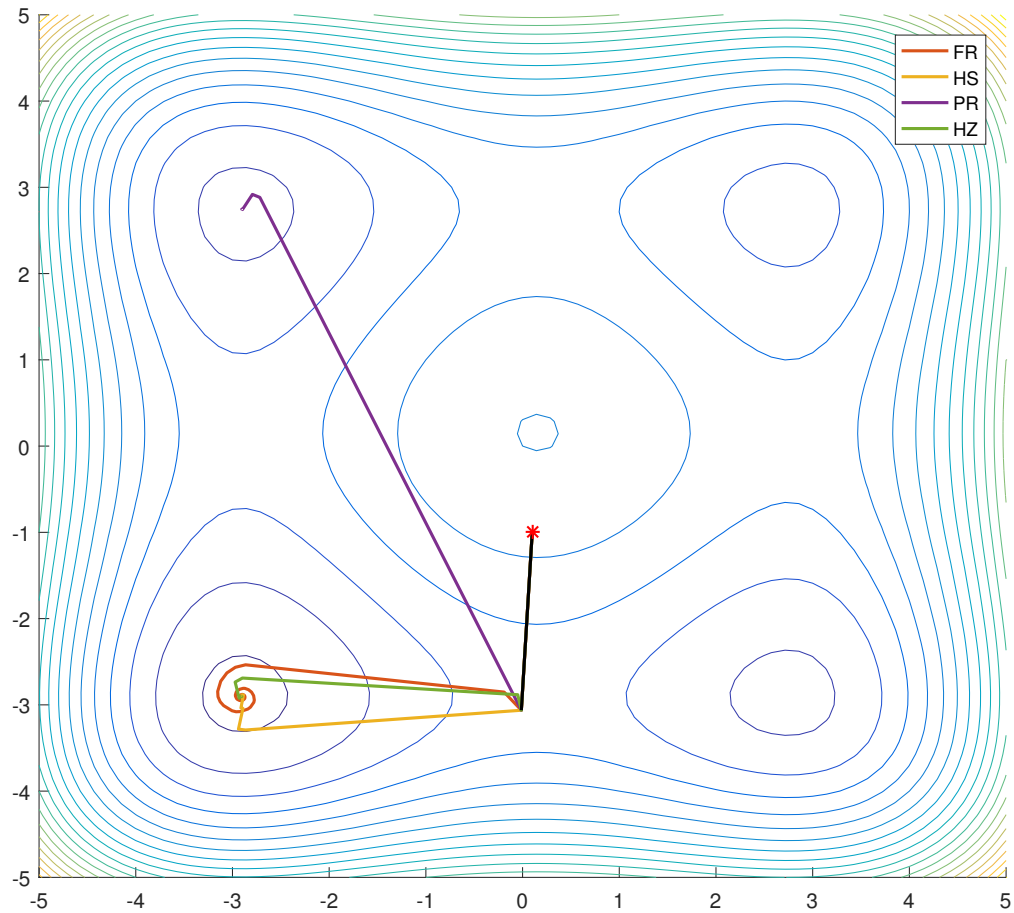


Рис. 3.2. Сравнение методов: начальное приближение (0.1, -1).

Метод	FR	HS	PR	HZ
Число итераций	239	19	12	17

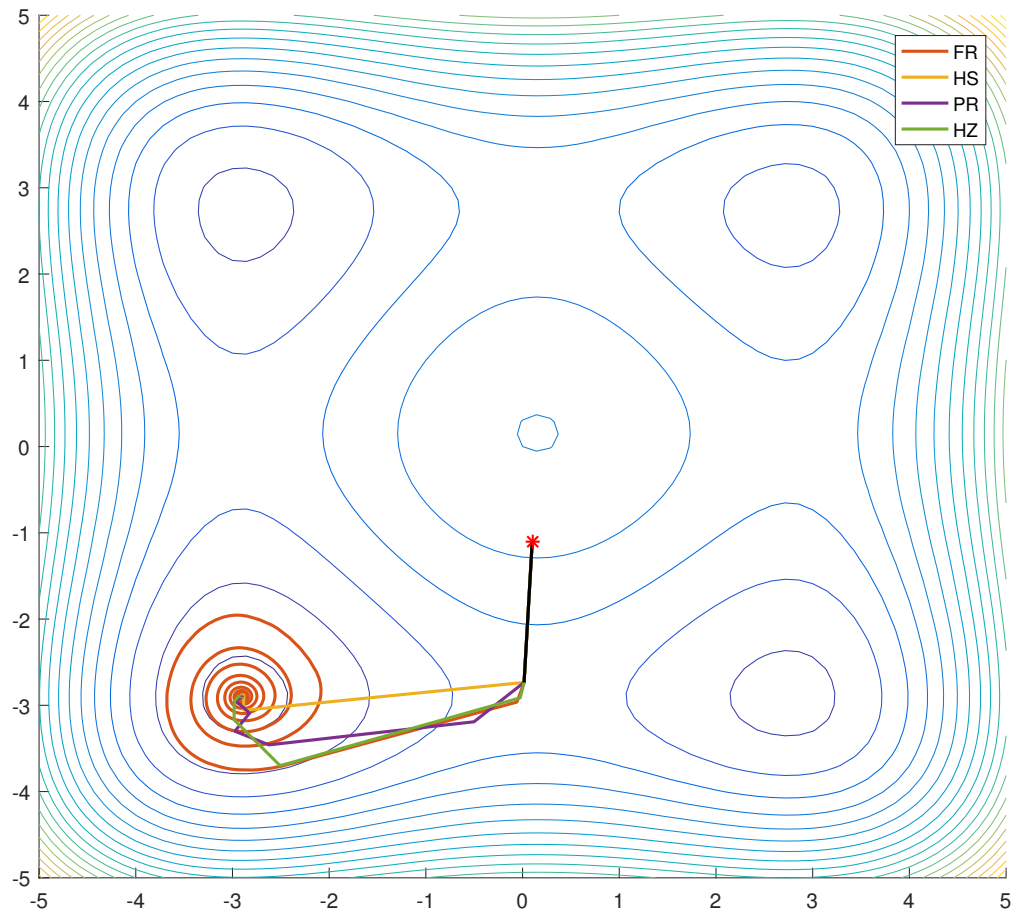


Рис. 3.3. Сравнение методов: начальное приближение $(0.1, -1.1)$.

Метод	FR	HS	PR	HZ
Число итераций	-	23	15	19

Метод Флетчера-Ривза не добился требуемой точности за 1000 итераций.

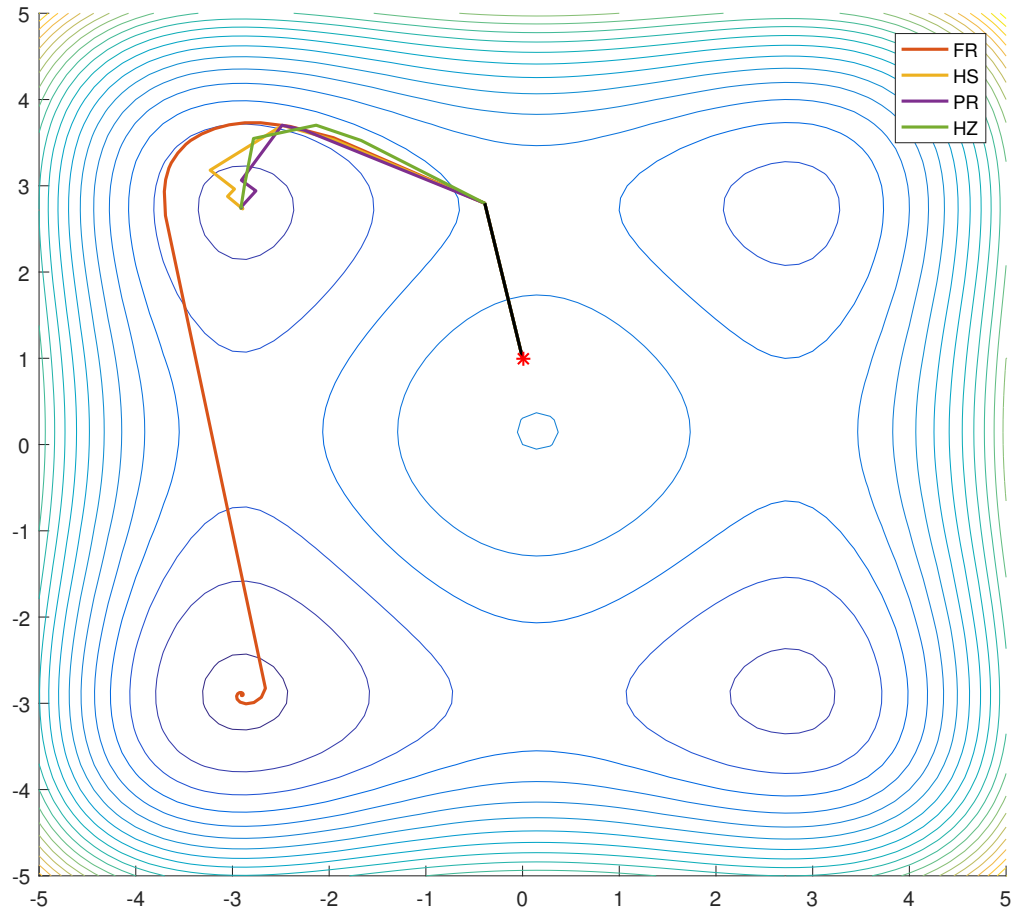


Рис. 3.4. Сравнение методов: начальное приближение $(0, -1)$.

Метод	FR	HS	PR	HZ
Число итераций	86	15	15	22

3.3. Анализ результатов

Как видно из полученных результатов, метод Флетчера-Ривза склонен делать большое количество итераций, незначительно уменьшая значение целевой функции.

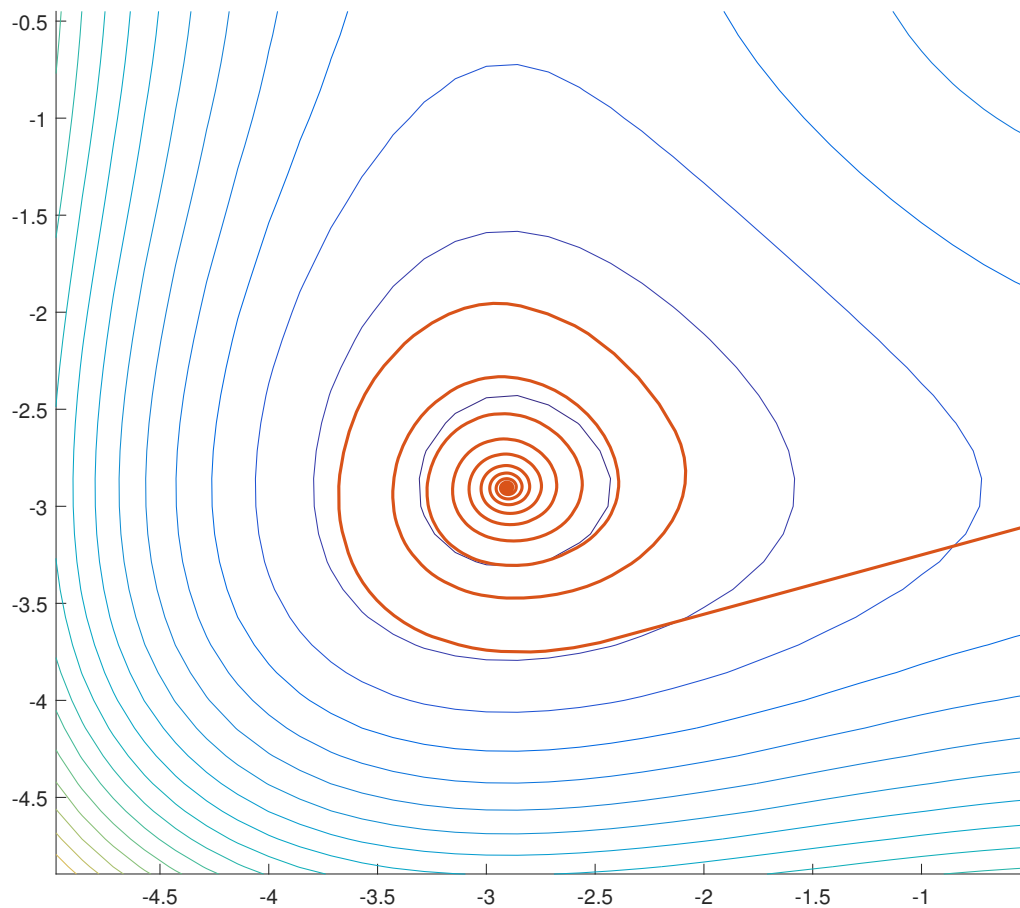
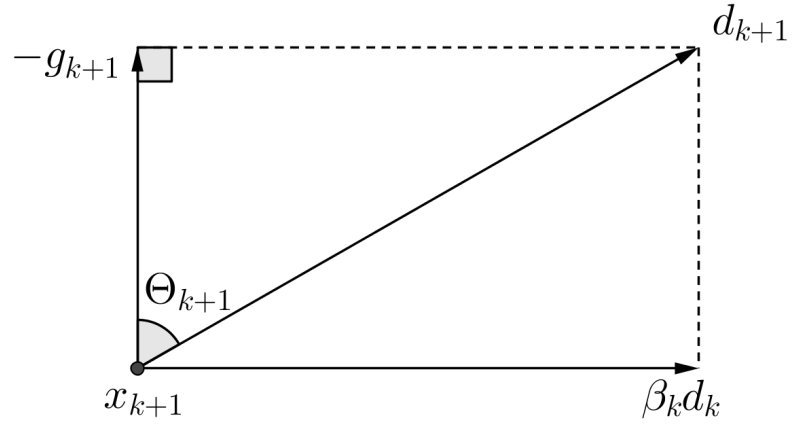


Рис. 3.5. Поведение метода Флетчера-Ривза.

В статье [5] Пауэлл следующим образом аргументировал такое поведение метода Флетчера-Ривза. Предположим, что используется точный линейный поиск (1.12). Следовательно, выполняется условие (1.13).



Тогда верны следующие равенства

$$\beta_k \|d_k\| = \operatorname{tg} \Theta_{k+1} \|g_{k+1}\|, \quad (3.2)$$

$$\|d_k\| = \sec \Theta_k \|g_k\|. \quad (3.3)$$

Выражая d_k из (3.2), (3.3) и используя определение β_k^{FR} , получим

$$\operatorname{tg} \Theta_{k+1} = \sec \Theta_k \frac{\|g_{k+1}\|}{\|g_k\|} > \operatorname{tg} \Theta_k \frac{\|g_{k+1}\|}{\|g_k\|}. \quad (3.4)$$

Если угол $\Theta_k \approx \frac{\pi}{2}$, то направление d_k почти перпендикулярно $-g_k$ и метод совершает совсем небольшой шаг в этом направлении. Тогда $\|g_{k+1} - g_k\| \approx 0$, а значит $\frac{\|g_{k+1}\|}{\|g_k\|} \approx 1$. Используя неравенство (3.4) приходим к выводу, что угол $\Theta_{k+1} \approx \frac{\pi}{2}$. Такая особенность метода делает его особенно требовательным к обновлению.

Глава 4

Минимизация циклической функции

Применим метод сопряженных градиентов для решения конкретной задачи.

4.1. Неравенство Шапиро

Рассмотрим циклическую функцию вида

$$F_n(x) = \frac{x_1}{x_2 + x_3} + \frac{x_2}{x_3 + x_4} + \dots + \frac{x_{n-1}}{x_n + x_1} + \frac{x_n}{x_1 + x_2}, \quad (4.1)$$

где $x_i \geq 0$, $i \in 1 : n$, и все знаменатели отличны от нуля.

В 1954 г. Г. Шапиро выдвинул гипотезу о том, что при $n \geq 3$ справедливо следующее неравенство:

$$F_n(x) \geq \frac{n}{2}. \quad (4.2)$$

На тот момент у него имелось доказательство только для случаев $n = 3$ и $n = 4$.

Отметим важное свойство этого неравенства:

Лемма 1. *Если неравенство Шапиро неверно при некотором $n = k$, то оно неверно и при $n = k + 2$.*

Доказательство. Пусть $F_k(x') < \frac{k}{2}$. Рассмотрим $x'' = (x'_1, \dots, x'_n, x'_1, x'_2)$. Заметим, что $F_{k+2}(x'') = F_k(x') + 1$, следовательно, $F_{k+2}(x'') < \frac{k+2}{2}$. \square

Гипотеза Шапиро вызвала широкий интерес математиков, но только к 1989 г. коллективными усилиями удалось установить, что неравенство Шапиро верно для четных $n \leq 12$ и нечетных $n \leq 23$. Для $n = 14$ и $n = 25$ неравенство нарушается, а значит, по доказанной лемме, нарушается и для всех четных $n > 14$ и нечетных $n > 25$.

4.2. Постановка задачи и метод решения

Задача заключалась в поиске значений x , при которых нарушается неравенство Шапиро в случаях $n = 14$ и $n = 25$. Для этого рассмотрим экстремальную задачу следующего вида:

$$F_n(x) \rightarrow \min, \quad x_i \geq 0, \quad i \in 1 : n. \quad (4.3)$$

Стоит отметить, что при $n \geq 3$ для всех векторов x с положительными компонентами справедлива оценка Дринфельда

$$F_n(x) > c \frac{n}{2}, \quad (4.4)$$

где $c = 0.989$ [6].

Сделаем замену $x_i = y_i^2$ и перейдем к задаче безусловной минимизации

$$G_n(y) = \frac{y_1^2}{y_2^2 + y_3^2} + \frac{y_2^2}{y_3^2 + y_4^2} + \dots + \frac{y_{n-1}^2}{y_n^2 + y_1^2} + \frac{y_n^2}{y_1^2 + y_2^2} \rightarrow \min_{y \in \mathbb{R}^n}. \quad (4.5)$$

Данная задача решалась методом Хестенса-Штифеля. Линейным поиск удовлетворял условию Армихо (1.14) с параметром $\mu = 0.5$. Обновление происходило каждые n итераций или при нарушении свойства убывания. Метод останавливался на итерации k , если выполнялось условие $\|g_k\| < 10^{-6}$.

4.3. Начальное приближение

Запуск алгоритма из случайных начальных приближений редко приводил к тому, что в получившихся точках значение функции было меньше $\frac{n}{2}$. Однако анализ успешных случаев при $n = 14$ позволил сделать предположение о структуре решения. На этом основании начальное приближение выбиралось следующим:

$$y_0 = [p, s_1, \dots, p, s_7],$$

где p — случайная величина, равномерно распределенная на отрезке $[0.7, 0.9]$, а s_i — независимые случайные величины, равномерно распределенные на отрезке $[0, 0.2]$. Алгоритм был запущен 10000 раз из точек такого вида и каждый раз получалась точка, значение функции в которой было меньше $\frac{n}{2}$.

При $n = 25$ успеха удалось добиться, когда в качестве начального приближения брались точки вида

$$y_0 = [p, s_1, \dots, p, s_{12}, p],$$

где p и s_i — такие же случайные величины, что и при $n = 14$. Алгоритм был снова 10000 раз запущен из точек указанного вида и вновь в каждом случае привел к решению. Стоит отметить, что значение функции в точках, взятых в качестве начального приближения, для $n = 14$ очень близко к $\frac{n}{2}$, а для $n = 25$ уже достаточно сильно отличается от $\frac{n}{2}$.

4.4. Результаты вычислений

Было получено большое количество значений x , при которых $F_n(x) < \frac{n}{2}$ для $n = 14$ и $n = 25$. Многие из них были приведены к целочисленному виду.

1. $n = 14$

- $x = [78, 7, 75, 8, 71, 6, 70, 3, 71, 1, 74, 1, 77, 3], \quad F_{14}(x) = 6.999996;$
- $x = [73, 8, 70, 6, 68, 3, 69, 1, 72, 1, 75, 3, 76, 7], \quad F_{14}(x) = 6.999994;$
- $x = [74, 6, 72, 3, 73, 1, 76, 1, 79, 3, 80, 7, 77, 8], \quad F_{14}(x) = 6.999991.$

2. $n = 25$

- $x = [39, 39, 38, 28, 36, 19, 35, 11, 35, 4, 39, 0, 46, 0, 55, 0, 65, 0, 77, 12, 79, 29, 68, 38, 53], \quad F_{25}(x) = 12.49896;$
- $x = [49, 48, 47, 35, 44, 23, 43, 13, 44, 5, 48, 0, 57, 0, 68, 0, 80, 0, 95, 15, 97, 36, 83, 47, 65], \quad F_{25}(x) = 12.49885;$
- $x = [43, 43, 41, 30, 39, 20, 38, 12, 39, 5, 43, 0, 51, 0, 60, 0, 71, 0, 84, 13, 86, 32, 74, 41, 57], \quad F_{25}(x) = 12.49881.$

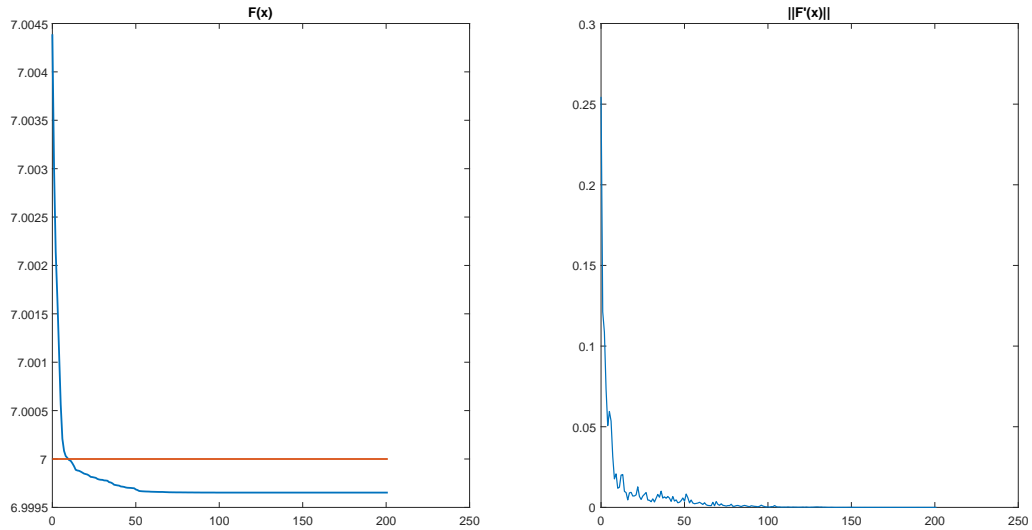


Рис. 4.1. Случай $n = 14$. Первое начальное приближение.

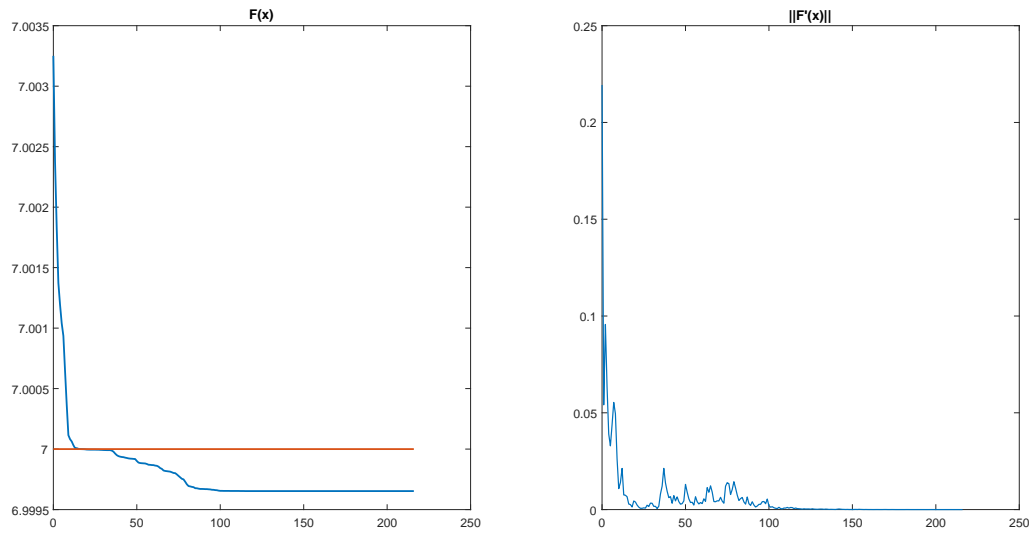


Рис. 4.2. Случай $n = 14$. Второе начальное приближение.

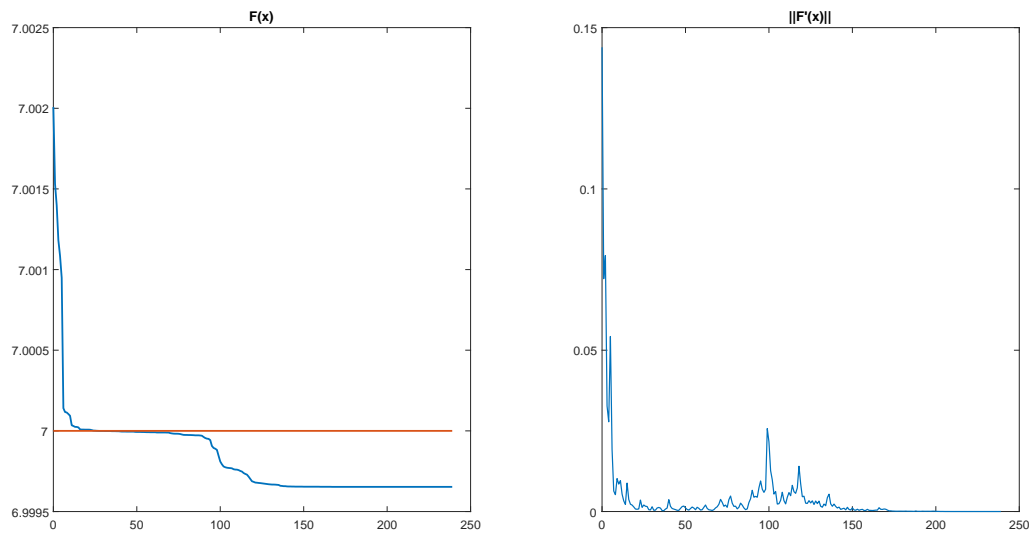
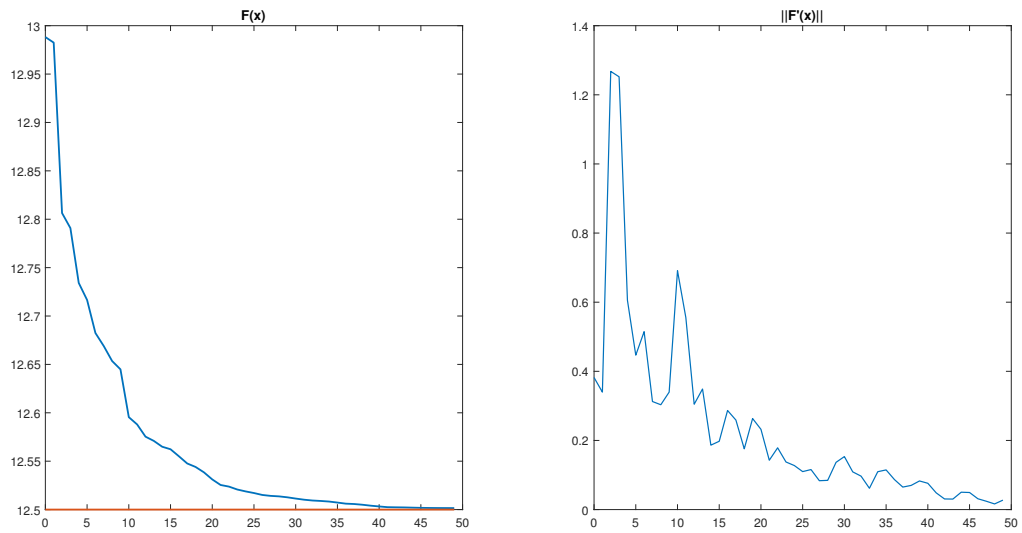
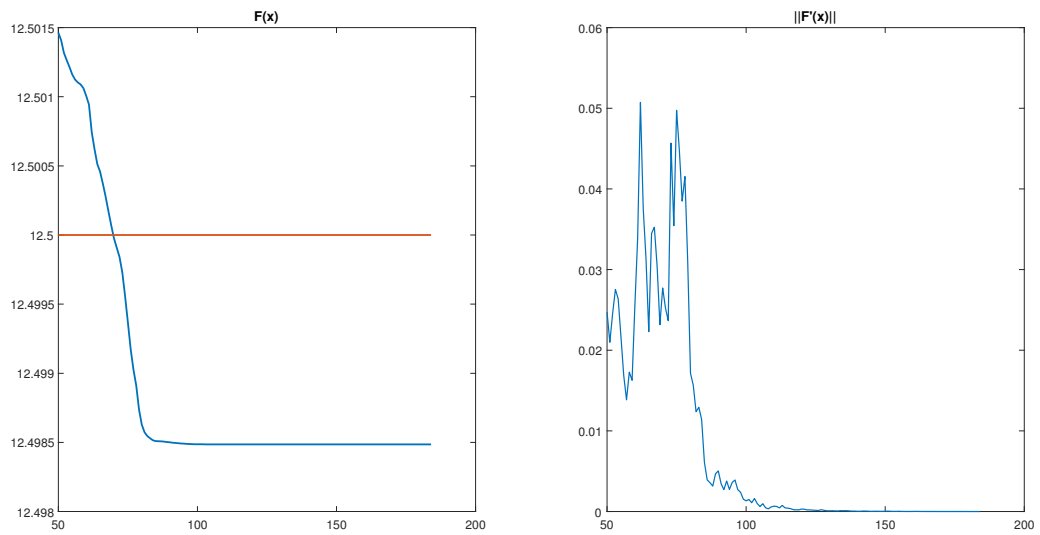


Рис. 4.3. Случай $n = 14$. Третье начальное приближение.

Рис. 4.4. Случай $n = 25$. До 50-й итерации.Рис. 4.5. Случай $n = 25$. После 50-й итерации.

Глава 5

Геометрический вариант метода сопряженных градиентов

5.1. Особенности геометрического варианта

Рассмотрим геометрический вариант метода сопряженных градиентов, описанный в [7], и применим его для решения задачи (1.6). В этом варианте направление d_k задается правилом

$$\begin{aligned} d_{k+1} &= (1 - \lambda_k) - g_{k+1} + \lambda_k d_k, \\ \lambda_k &= \frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2 + \|d_k\|^2} \end{aligned} \quad (5.1)$$

Метод имеет геометрическую интерпретацию. Пусть треугольник, натянутый на вектора $-g_{k+1}$, d_k , невырожден. Рассмотрим треугольники $\triangle ABD$ и $\triangle ACD$, изображенные на рис. 5.1. Тогда

$$\frac{S_{\triangle ABD}}{S_{\triangle ACD}} = \frac{\|g_{k+1}\|^2}{\|d_k\|^2}. \quad (5.2)$$

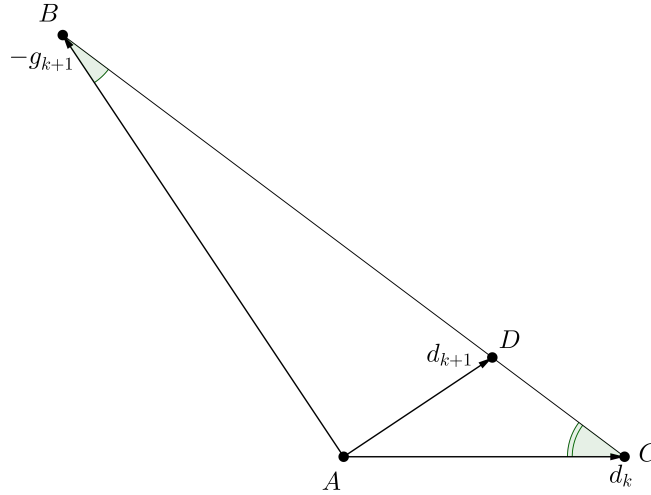


Рис. 5.1. Геометрический вариант метода.

Доказательство. Используя формулу (5.1) и теорему синусов, получим

$$\frac{S_{\triangle ABD}}{S_{\triangle ACD}} = \frac{BA}{CA} \frac{BD}{CD} \frac{\sin \angle ABD}{\sin \angle ACD} = \frac{\|g_{k+1}\|}{\|d_k\|} \frac{\|g_{k+1}\|^2}{\|d_k\|^2} \frac{\|d_k\|}{\|g_{k+1}\|} = \frac{\|g_{k+1}\|^2}{\|d_k\|^2}. \quad (5.3)$$

□

5.2. Сходимость

Заметим, что формула (5.1), как и (1.9), гарантирует выполнение свойства убывания для $k \geq 1$, если используется точный поиск (1.12). Данный факт позволяет использовать Замечание 1 для доказательства сходимости геометрического варианта метода сопряженных градиентов при точном поиске (1.12).

Теорема 3. Пусть функция f удовлетворяет Условиям 1, 2. Если в геометрическом варианте метода сопряженных градиентов значение α_k вычисляется точным линейным поиском (1.12), то

$$\lim_{k \rightarrow \infty} \|g_k\| = 0. \quad (5.4)$$

Доказательство. Пусть утверждение (5.4) неверно, тогда существует константа $\varepsilon > 0$ такая, что

$$\|g_k\|^2 \geq \varepsilon > 0 \quad \forall k \geq 1. \quad (5.5)$$

Воспользуемся Замечанием 1

$$f(x_k) - f(x_{k+1}) \geq c \cos^2 \Theta_k \|g_k\|^2, \quad (5.6)$$

где $c > 0$. Из формулы (5.1) и условия (1.13) получаем

$$\cos^2 \Theta_k \|g_k\|^2 = \|d_k\|^2. \quad (5.7)$$

Тогда из (5.6) следует

$$f(x_k) - f(x_{k+1}) \geq c \|d_k\|^2. \quad (5.8)$$

Докажем по индукции, что

$$\|d_k\|^2 \geq \frac{\varepsilon}{k} \quad \forall k \geq 1. \quad (5.9)$$

При $k = 1$ неравенство верно. Сделаем индукционный переход от k к $k + 1$. Воспользовавшись формулой (5.1) и условием (1.13), получим выражение для $\|d_{k+1}\|^2$

$$\|d_{k+1}\|^2 = \frac{\|g_{k+1}\|^2 \|d_k\|^2}{\|g_{k+1}\|^2 + \|d_k\|^2} = \frac{1}{\frac{1}{\|d_k\|^2} + \frac{1}{\|g_{k+1}\|^2}} \quad (5.10)$$

Далее, в силу индукционного предположения и утверждения (5.5), имеем

$$\frac{1}{\frac{1}{\|d_k\|^2} + \frac{1}{\|g_{k+1}\|^2}} \geq \frac{1}{\frac{k}{\varepsilon} + \frac{1}{\varepsilon}} = \frac{\varepsilon}{k + 1}, \quad (5.11)$$

а значит

$$\|d_{k+1}\|^2 \geq \frac{\varepsilon}{k + 1}. \quad (5.12)$$

Возвращаясь к (5.8), получим

$$f(x_k) - f(x_{k+1}) \geq c \frac{\varepsilon}{k} \quad \forall k \geq 1, \quad (5.13)$$

из чего следует

$$\lim_{k \rightarrow \infty} f(x_k) = -\infty, \quad (5.14)$$

что противоречит условию ограниченности функции f снизу. \square

Если же линейный поиск выполняется неточно, то выполнение свойства убывания для всех $k \geq 1$ требует доказательства.

Лемма 2. Пусть вектора $-g_{k+1} \neq \mathbb{O}$, $d_k \neq \mathbb{O}$ такие, что $|\langle -g_{k+1}, d_k \rangle| \leq c \|d_k\|^2$, где $c \in (0, 1)$. Тогда $d_{k+1} \neq \mathbb{O}$ и выполняется неравенство

$$\frac{\langle -g_{k+1}, d_{k+1} \rangle}{\|d_{k+1}\|^2} \leq \frac{c^2 + 1}{1 - c}. \quad (5.15)$$

Доказательство. Из формулы (5.1) находим

$$\begin{aligned} \langle -g_{k+1}, d_{k+1} \rangle &= (1 - \lambda_k) \|g_{k+1}\|^2 + \lambda_k \langle -g_{k+1}, d_k \rangle = \\ &= \lambda_k (\|d_k\|^2 + \langle -g_{k+1}, d_k \rangle). \end{aligned} \quad (5.16)$$

$$\begin{aligned} \|d_{k+1}\|^2 &= (1 - \lambda_k)^2 \|g_{k+1}\|^2 + \lambda_k^2 \|d_k\|^2 + 2\lambda_k(1 - \lambda_k) \langle -g_{k+1}, d_k \rangle = \\ &= \lambda_k (\|d_k\|^2 + 2(1 - \lambda_k) \langle -g_{k+1}, d_k \rangle). \end{aligned} \quad (5.17)$$

Покажем, что $\|d_{k+1}\|^2 > 0$. Пусть, напротив, $\|d_{k+1}\|^2 = 0$, тогда

$$\|d_k\|^2 + 2(1 - \lambda_k) \langle -g_{k+1}, d_k \rangle = 0. \quad (5.18)$$

Поделив на $\|d_k\|^2 > 0$, получим

$$1 + \frac{2\langle -g_{k+1}, d_k \rangle}{\|g_{k+1}\|^2 + \|d_k\|^2} = 0, \quad (5.19)$$

что равносильно

$$2\langle g_{k+1}, d_k \rangle = \|g_{k+1}\|^2 + \|d_k\|^2. \quad (5.20)$$

Но такое равенство возможно лишь в случае $g_{k+1} = d_k$, что противоречит условию

$$|\langle -g_{k+1}, d_k \rangle| \leq c \|d_k\|^2. \quad (5.21)$$

Используя выражения (5.16), (5.17), получим

$$\begin{aligned} \frac{\langle -g_{k+1}, d_{k+1} \rangle}{\|d_{k+1}\|^2} &= \frac{\|d_k\|^2 + \langle -g_{k+1}, d_k \rangle}{\|d_k\|^2 + 2(1 - \lambda_k) \langle -g_{k+1}, d_k \rangle} = \\ &= \frac{\|d_k\|^2 + \langle -g_{k+1}, d_k \rangle}{\|d_k\|^2 + 2 \frac{\|d_k\|^2}{\|g_{k+1}\|^2 + \|d_k\|^2} \langle -g_{k+1}, d_k \rangle}. \end{aligned} \quad (5.22)$$

Пусть $|\langle -g_{k+1}, d_k \rangle| = c_1 \|d_k\|^2$, где $c_1 \in [0, c]$. Рассмотрим два случая.

1. $\langle -g_{k+1}, d_k \rangle = c_1 \|d_k\|^2 \geq 0$. Используя выражение (5.22), получим

$$\frac{\langle -g_{k+1}, d_{k+1} \rangle}{\|d_{k+1}\|^2} \leq \frac{\|d_k\|^2 + c_1 \|d_k\|^2}{\|d_k\|^2} = 1 + c_1 \leq 1 + c. \quad (5.23)$$

2. $\langle -g_{k+1}, d_k \rangle = -c_1 \|d_k\|^2 < 0$. Используя выражение (5.22), получим

$$\begin{aligned} \frac{\langle -g_{k+1}, d_{k+1} \rangle}{\|d_{k+1}\|^2} &\leq \frac{\|d_k\|^2 - c_1 \|d_k\|^2}{\|d_k\|^2 - 2 \frac{\|d_k\|^2}{c_1^2 \|d_k\|^2 + \|d_k\|^2} c_1 \|d_k\|^2} = \\ &= \frac{1 - c_1}{1 - \frac{2c_1}{c_1^2 + 1}} = \frac{c_1^2 + 1}{1 - c_1} \leq \frac{c^2 + 1}{1 - c}. \end{aligned} \quad (5.24)$$

Осталось заметить, что

$$1 + c \leq \frac{c^2 + 1}{1 - c}. \quad (5.25)$$

Равенство в (5.15) достигается, когда $g_{k+1} = cd_k$. \square

Теорема 4. Зафиксируем $c \in (0, 1)$. Пусть в геометрическом варианте метода сопряженных градиентов значение α_k удовлетворяет условию (1.19) с параметром $\nu \leq \frac{c(1-c)}{c^2+1}$. Тогда свойство убывания выполняется для всех $k \geq 1$.

Доказательство. Докажем по индукции, что неравенство

$$0 < \langle -g_k, d_k \rangle \leq \frac{c^2 + 1}{1 - c} \|d_k\|^2. \quad (5.26)$$

выполняется при всех $k \geq 1$. При $k = 1$ неравенство верно. Сделаем индукционный переход от k к $k + 1$. Из условия (1.19) и индукционного предположения получим

$$|\langle -g_{k+1}, d_k \rangle| \leq \nu \langle -g_k, d_k \rangle \leq \frac{c(1-c)}{c^2+1} \frac{c^2+1}{1-c} \|d_k\|^2 = c \|d_k\|^2. \quad (5.27)$$

Из полученного неравенства и выражения (5.16) получим

$$0 < \langle -g_{k+1}, d_{k+1} \rangle, \quad (5.28)$$

а из Леммы 2

$$\langle -g_{k+1}, d_{k+1} \rangle \leq \frac{c^2 + 1}{1 - c} \|d_{k+1}\|^2. \quad (5.29)$$

\square

Следствие 2. Пусть в методе сопряженных градиентов d_{k+1} вычисляется по формуле (5.1), а значение α_k удовлетворяет условию (1.19) с параметром $\nu \leq \frac{1}{2}(\sqrt{2} - 1)$. Тогда свойство убывания выполняется для всех $k \geq 1$.

Доказательство. Максимум функции $\frac{c(1-c)}{c^2+1}$, $c \in (0, 1)$ равен $\frac{1}{2}(\sqrt{2} - 1)$ и достигается при $c = \sqrt{2} - 1$. Воспользовавшись Теоремой 4, получим требуемый результат. \square

Заключение

1. В рамках данной работы проведен обзор и численное сравнение различных вариантов метода сопряженных градиентов.
2. С помощью метода сопряженных градиентов решена задача о построении контр-примеров для неравенства Шапиро.
3. Доказана теорема о сходимости геометрического варианта метода сопряженных градиентов при точном линейном поиске.
4. Установлено свойство убывания целевой функции в геометрическом варианте метода сопряженных градиентов при неточном линейном поиске.

Предварительные результаты опубликованы в [8].

Список литературы

1. Hestenes M. R., Stiefel E. Method of Conjugate Gradients for Solving Linear Systems // Journal of Research of the National Bureau of Standards. 1952. Vol. 49. P. 409–436.
2. Малозёмов В. Н. О методе сопряжённых градиентов. Семинар «DHA & CAGD». Избранные доклады. 28.04.2012. URL: <http://dha.spb.ru/rep12.shtml#0428>.
3. Nocedal J., Wright S. J. Numerical Optimization. 2nd edition. Springer, 2006. P. 664.
4. Pytlak R. Conjugate Gradient Algorithms in Nonconvex Optimization. Springer, 2009. P. 478.
5. Powell M. J. D. Restart Procedures of the Conjugate Gradient Method // Mathematical Programming. 1977. Vol. 12. P. 241–254.
6. Малозёмов В. Н. Циклические функции и экстремальные задачи. Семинар «CNSA & NDO». Избранные доклады. 27.08.2015. URL: <http://www.apmath.spbu.ru/cnsa/rep15.shtml#0827>.
7. Малозёмов В. Н. Варианты метода сопряжённых градиентов. Семинар «CNSA & NDO». Избранные доклады. 29.10.2015. URL: <http://www.apmath.spbu.ru/cnsa/rep15.shtml#1029>.
8. Плоткин А. В. Минимизация циклической функции // Процессы управления и устойчивость. Vol. 3 (19). 2016. (принята в печать).
9. Fletcher R., Reeves C. M. Function Minimization by Conjugate Gradients // Computer Journal. 1964. Vol. 7. P. 149–154.
10. Golub G. H., O’Leary D. P. Some History of the Conjugate Gradient Methods and Lanczos Algorithms: 1948–1976 // SIAM Rev. 1989. Vol. 31. P. 50–102.
11. Hager W. W., Zhang H. A Survey of Nonlinear Conjugate Gradient Methods // Pacific Journal of Optimization. 2006. Vol. 2. P. 335–358.
12. Dai Y. H. Nonlinear Conjugate Gradient Methods. Wiley Encyclopedia of Operations Research and Management Science. 15.02.2011.